

CLAIMS

- 1 1. A method for enabling recovery from two or fewer concurrent failures of storage de-
 2 vices in a storage array, the method comprising the steps of:
 3 providing the array with a predetermined number of storage devices, including a
 4 plurality of first devices configured to store data and row parity, and one diagonal parity
 5 device configured to store diagonal parity, wherein the predetermined number of storage
 6 devices n is $p+1$ and wherein p is a prime number;
 7 dividing each device into blocks;
 8 organizing the blocks into stripes that contain a same number of blocks in each
 9 device, wherein each stripe comprises $n-2$ rows of blocks;
 10 defining the diagonal parity along diagonal parity sets that span the first devices,
 11 wherein the diagonal parity sets wrap around within a group of $n-2$ rows so that all blocks
 12 belonging to diagonal parity sets of a stripe are stored in the stripe; and
 13 computing and storing the diagonal parity for all of the diagonal parity sets except
 14 one on the diagonal parity device.
- 1 2. The method of Claim 1 wherein row parity blocks in a stripe are all stored on a single
 2 device.
- 1 3. The method of Claim 1 wherein the first devices include a plurality of data devices,
 2 and wherein the data devices are not all present and absent data devices are treated as
 3 containing zero-valued data.
- 1 4. The method of Claim 1 wherein computation of row parity blocks is replaced by a re-
 2 dundant encoding of data blocks in a stripe that is tolerant of single device failures.
- 1 5. The method of Claim 1 wherein a total length of the devices is unequal and wherein
 2 each row has one row parity block and one diagonal parity block.

1 6. The method of Claim 1 wherein locations of parity blocks shift from device to device
2 within different stripes.

1 7. A system configured to enable recovery from two or fewer concurrent failures of two
2 storage devices, the system comprising:

3 an array having a predetermined number of storage devices, including a plurality
4 of first devices configured to store data and row parity, and one diagonal parity device
5 configured to store diagonal parity, wherein the predetermined number of storage devices
6 n is $p+1$ and wherein p is a prime number;

7 a storage operating system including a device storage layer configured to imple-
8 ment a row-diagonal (R-D) parity technique that computes the diagonal parity along di-
9 agonal parity sets that span the first devices, and that stores the diagonal parity for all of
10 the diagonal parity sets except one on the diagonal parity device; and

11 a processing element configured to execute the storage operating system to
12 thereby invoke storage access operations to and from the array in accordance with the R-
13 D parity technique.

1 8. The system of Claim 7 wherein row parity blocks in a stripe are all stored on a single
2 device.

1 9. The system of Claim 7 wherein computation of row parity blocks is replaced by a re-
2 dundant encoding of data blocks in a stripe that is tolerant of single device failures.

1 10. The system of Claim 7 wherein the device storage layer is a RAID layer and wherein
2 the storage devices are disks.

1 11. The system of Claim 10 wherein the RAID layer further divides each disk into
2 blocks and organizes the blocks into stripes.

- 1 12. The system of Claim 11 wherein each stripe comprises $n-2$ rows of blocks, with
2 $n=p+1$, wherein each row contains one block from each disk.
- 1 13. The system of Claim 12 wherein a total length of the devices is unequal and wherein
2 each row has one row parity block and one diagonal parity block.
- 1 14. The system of Claim 10 wherein the RAID layer is implemented in logic circuitry.
- 1 15. The system of Claim 7 wherein the devices are one of video tape, magnetic tape, op-
2 tical, DVD, bubble memory, magnetic disk, electronic random access memory and mi-
3 cro-electro mechanical storage devices.
- 1 16. The system of Claim 7 wherein the devices are media adapted to store information
2 contained within the data and parity blocks.
- 1 17. The system of Claim 7 wherein the first devices include a plurality of data devices,
2 and wherein the data devices are not all present and absent data devices are treated as
3 containing zero-valued data.
- 1 18. Apparatus for enabling recovery from two or fewer concurrent failures of two stor-
2 age devices in a storage array, the apparatus comprising:
3 means for providing the array with a predetermined number of storage devices,
4 including a plurality of first devices configured to store data and row parity, and one di-
5 agonal parity device configured to store diagonal parity, wherein the predetermined num-
6 ber of storage devices n is $p+1$ and wherein p is a prime number;
7 means for computing the diagonal parity along diagonal parity sets that span the
8 first devices; and
9 means for storing the diagonal parity for all of the diagonal parity sets except one
10 on the diagonal parity device.

1 19. The apparatus of Claim 18 wherein row parity blocks in a stripe are all stored on a
2 single device.

1 20. The system of Claim 18 wherein computation of row parity blocks is replaced by a
2 redundant encoding of data blocks in a stripe that is tolerant of single device failures.

1 21. The apparatus of Claim 18 further comprising:
2 means for dividing each device into blocks; and
3 means for organizing the blocks into stripes.

1 22. The apparatus of Claim 21 wherein the means for storing comprises means for stor-
2 ing diagonal parity blocks on the diagonal parity disk for each of the diagonal parity sets
3 of a stripe except one.

1 23. The apparatus of Claim 18 wherein the first devices include a plurality of data de-
2 vices, and wherein the data devices are not all present and absent data devices are treated
3 as containing zero-valued data.

1 24. A computer readable medium containing executable program instructions for ena-
2 bling recovery from two or fewer concurrent failures of two storage devices in a storage
3 array, the executable program instructions comprising program instructions for:

4 providing the array with a predetermined number of storage devices, including a
5 plurality of first devices configured to store data and row parity, and one diagonal parity
6 device configured to store diagonal parity, wherein the predetermined number of storage
7 devices n is $p+1$ and wherein p is a prime number;

8 computing the diagonal parity along diagonal parity sets that span the first de-
9 vices; and

10 storing the diagonal parity for all of the diagonal parity sets except one on the di-
11 agonal parity device.

1 25. The computer readable medium of Claim 24 wherein row parity blocks in a stripe are
2 all stored on a single device.

1 26. The computer readable medium of Claim 24 wherein computation of row parity
2 blocks is replaced by a redundant encoding of data blocks in a stripe that is tolerant of
3 single device failures.

1 27. The computer readable medium of Claim 24 further comprising program instructions
2 for:

3 dividing each device into blocks; and
4 organizing the blocks into stripes.

1 28. The computer readable medium of Claim 27 wherein the program instruction for
2 storing comprises program instructions for storing diagonal parity blocks on the diagonal
3 parity disk for each of the diagonal parity sets of a stripe except one.

1 29. The computer readable medium of Claim 24 wherein the first devices include a plu-
2 rality of data devices, and wherein the data devices are not all present and absent data de-
3 vices are treated as containing zero-valued data.

1 30. A method for enabling recovery from two or fewer concurrent failures of two storage
2 devices in a storage array, the method comprising the steps of:

3 providing the array with a predetermined number of storage devices equal to n ,
4 where $n=p+1$ and p is a prime number, including $n-2$ data devices configured to hold
5 data, one row parity device and one diagonal parity device;

6 dividing each device into blocks of fixed size;

7 organizing the blocks into stripes, wherein each stripe comprises $n-2$ rows of
8 blocks such that each device contains $n-2$ blocks per stripe;

9 computing the diagonal parity along $n-2$ of $n-1$ parity sets (diagonal) of blocks
10 that span the data devices and the row parity device, wherein the blocks of a diagonal are

11 contained within a stripe, each diagonal containing a set of $n-2$ blocks selected from
 12 among data and row parity blocks in a stripe such that no diagonal contains more than
 13 one block from a same row, each diagonal further containing $n-2$ blocks from a group of
 14 $n-2$ rows and no diagonal containing two blocks from a same device;

15 storing, on the row parity device, row parity for rows of blocks computed across
 16 all of the data devices such that each row contains one block from each data device and
 17 no row contains two blocks from a same diagonal; and

18 storing, in the $n-2$ blocks in the stripe on the diagonal parity device, the diagonal
 19 parity of $n-2$ of the $n-1$ diagonals, wherein selection of the diagonals stored on the diago-
 20 nal parity device is predetermined and wherein an order of placement of the $n-2$ diagonal
 21 parity blocks in the stripe on the diagonal parity device is predetermined.

1 31. The method of Claim 30 wherein the rows containing each stripe are contiguous.

1 32. The method of Claim 31 wherein the diagonals are formed by numbering the data,
 2 row parity and diagonal parity devices from 0 to $n-1$, and numbering the rows from 0 to
 3 $n-3$, then assigning the block at device i , row j to diagonal $(i+j) \bmod (n-2)$.

1 33. The method of Claim 30 wherein the data devices are not all present and wherein ab-
 2 sent data devices are treated as containing zero-valued data.

1 34. The method of Claim 30 wherein a total length of the devices is unequal and wherein
 2 each row has one row parity block and one diagonal parity block.

1 35. The method of Claim 30 wherein the storage devices are one of video tape, magnetic
 2 tape, optical, DVD, bubble memory, electronic random access memory, micro-electro
 3 mechanical and magnetic disk devices, and wherein the magnetic disk devices are one of
 4 hard disk drives and direct access storage devices.

1 36. A method for enabling recovery from two or fewer concurrent failures of two storage
2 devices in a storage array, the method comprising the steps of:

3 providing the array with a predetermined number of storage devices, including a
4 plurality of first devices configured to store data and row parity, and one diagonal parity
5 device configured to store diagonal parity;

6 computing the diagonal parity along diagonal parity sets that span the first de-
7 vices; and

8 storing the diagonal parity for all of the diagonal parity sets except one on the di-
9 agonal parity device.

1 37. The method of Claim 36 further comprising the steps of:

2 dividing each device into blocks; and

3 organizing the blocks into stripes.

1 38. The method of Claim 37 wherein the step of storing comprises the step of storing di-
2 agonal parity blocks on the diagonal parity disk for each of the diagonal parity sets of a
3 stripe except one.

1 39. The method of Claim 36 wherein the first devices include a plurality of data devices,
2 and wherein the data devices are not all present and absent data devices are treated as
3 containing zero-valued data.

1 40. The method of Claim 36 wherein the predetermined number of storage devices is n ,
2 and wherein $n=p+1$ and p is a prime number.

1 41. A method for enabling recovery from two or fewer concurrent failures of two storage
2 devices in a storage array, the method comprising the steps of:

3 providing the array with a predetermined number of storage devices equal to n ,
4 where $n=p+1$ and p is a prime number, including $n-2$ data devices configured to hold
5 data, one row parity device and one diagonal parity device;

6 dividing each device into blocks of fixed size;

7 organizing the blocks into stripes, wherein each stripe comprises $n-2$ rows of

8 blocks such that each device contains $n-2$ blocks per stripe and wherein locations of par-

9 ity blocks shift from device to device within different stripes;

10 computing the diagonal parity along $n-2$ of $n-1$ parity sets (diagonal) of blocks

11 that span the data devices and the row parity device, wherein the blocks of a diagonal are

12 contained within a stripe, each diagonal containing a set of $n-2$ blocks selected from

13 among data and row parity blocks in a stripe such that no diagonal contains more than

14 one block from a same row, each diagonal further containing $n-2$ blocks from a group of

15 $n-2$ rows and no diagonal containing two blocks from a same device;

16 storing, on the row parity device, row parity for rows of blocks computed across

17 all of the data devices such that each row contains one block from each data device and

18 no row contains two blocks from a same diagonal; and

19 storing, in the $n-2$ blocks in the stripe on the diagonal parity device, the diagonal

20 parity of $n-2$ of the $n-1$ diagonals, wherein selection of the diagonals stored on the diago-

21 nal parity device is predetermined and wherein an order of placement of the $n-2$ diagonal

22 parity blocks in the stripe on the diagonal parity device is predetermined.